International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 13, Issue 5, May-2025, Available online at: www.ijaresm.com

Hybrid Whole-Genome Assembly of Streptococcus Equi Subsp. Zooepidemicus MTCC3523 Utilizing Oxford Nanopore and Illumina Technologies for Functional Genomic Understanding of Enhanced Hyaluronic Acid Synthesis

Fatemeh Khatami Tousi[1], Seyedeh Zoha Tabatabaei[2]

----------**********----------

ABSTRACT

Recent advancements in sequencing technologies have transformed bacterial genomics by allowing for detailed genome characterization. Whole-genome sequencing (WGS) has become a cornerstone of bacterial taxonomy, functional genomics, and evolutionary studies. The development of hybrid sequencing methods that combine the precision of short-read technologies with the long-read capabilities of systems like Oxford Nanopore has significantly improved genome assembly and annotation, particularly for intricate and repetitive bacterial genomes. In this research, we provide a comparative evaluation of Illumina and Oxford Nanopore sequencing in the WGS of Streptococcus equi subsp. zooepidemicus MTCC 3523. While Illumina sequencing delivers highly accurate short reads, Nanopore technology offers long reads with relatively higher error rates. By merging both datasets, a hybrid assembly approach was utilized to enhance sequencing accuracy, gene prediction, and structural variant detection. The hybrid assembly yielded a 2.1 Mb genome comprised of 11 contigs, demonstrating greater contiguity and completeness. Gene prediction and functional annotation identified 2,008 coding sequences, which were associated with key metabolic pathways. A comparative genome analysis revealed nearly 100% similarity to the reference strain S. equi subsp. zooepidemicus NCTC 7023 and identified six major genomic rearrangements. Moreover, this study provides insights into the genetic mechanisms involved in the biosynthesis of hyaluronic acid (HA), a substance of considerable biomedical and industrial importance. The accurate identification of the hyaluronidase gene creates potential for targeted genetic modifications aimed at increasing HA production. Overall, this research emphasizes the combined advantages of utilizing both Illumina and Nanopore sequencing in microbial genomics. The hybrid method not only enhances the quality of genome assembly but also offers potential for strain improvement in large-scale HA production.

Keywords: Streptococcus Equi Subsp. Zooepidemicus MTCC 3523; Hyaluronic Acid; Oxford Nanopore; Illumina; Whole-Genome Sequencing

INTRODUCTION

The development of sequencing technologies has significantly changed the landscape of genomics, allowing for accurate and thorough characterization of bacterial genomes. Whole-genome sequencing (WGS) is an essential tool for exploring bacterial classification, functional genomics, and evolutionary connections. Hybrid sequencing methods, which combine short-read and long-read sequencing technologies, have emerged as a groundbreaking approach for genome assembly and annotation, especially for organisms with intricate, repetitive, or previously unsequenced genomes.

Oxford Nanopore and Illumina sequencing are two of the most commonly used platforms for hybrid genome assembly. Illumina sequencing, which is a second-generation technology, provides high-throughput short reads with outstanding accuracy, making it suitable for identifying single-nucleotide polymorphisms (SNPs), as well as small insertions and deletions. However, its limited read lengths hinder its ability to resolve repetitive areas and structural variations. On the other hand, Oxford Nanopore sequencing, classified as a fourth-generation technology, offers real-time long-read sequencing, which allows for the resolution of intricate genomic regions, structural variants, and entire plasmid structures. Nonetheless, Nanopore sequencing has a greater inherent error rate, which calls for hybrid methods that integrate the precision of short reads with the longer read lengths offered by long-read technologies.

Utilizing the benefits of both technologies, hybrid de novo assembly aids in the accurate and complete reconstruction of continuous bacterial genomes. This method is especially crucial for sequencing bacterial species that have uncertain or significantly variable genomic structures, enabling researchers to produce nearly complete assemblies that accurately depict the actual genetic framework of an organism(5). One of the most thoroughly examined bacterial strains for industrial uses is Streptococcus equi subsp. zooepidemicus MTCC 3523. This strain is well-known for its remarkable capacity to produce high-molecular-weight hyaluronic acid (HA), a polysaccharide that has important uses in medicine,

cosmetics, and biotechnology. HA is vital for tissue moisture retention, healing of wounds, and lubrication of joints, making it a critical component in skin treatments, therapies for osteoarthritis, eye surgery, and sophisticated drug delivery systems. Large-sized HA, characterized by its higher molecular weight, is especially prized for its improved viscoelastic characteristics, extended retention in biological tissues, and enhanced effectiveness in medical and cosmetic applications. The molecular weight of HA is a key factor in determining its biological functions, with high-molecular-weight HA demonstrating anti-inflammatory effects, immunomodulatory properties, and the ability to protect the extracellular matrix.

Although HA production in S. equi subsp. zooepidemicus is crucial for industry, it is impeded by hyaluronidase, an enzyme that catalyzes the breakdown of HA into smaller molecular fragments. The activity of hyaluronidase not only decreases the yield of HA but also undermines the quality of the final product by creating low-viscosity and biologically unstable HA fractions. To enhance HA production for industrial purposes, it is vital to eliminate or inhibit hyaluronidase activity, thus ensuring the preservation of high-molecular-weight HA and optimizing both product stability and effectiveness.

The management of hyaluronidase expression and activity continues to pose a significant challenge in the production of microbial hyaluronic acid. Techniques in genetic engineering designed to knock out or reduce the expression of the hyaluronidase gene require a detailed understanding of its location in the genome and its regulatory elements. Nevertheless, previous research has indicated variations in the genomic structure of S. equi subsp. zooepidemicus strains, making it essential to conduct strain-specific genomic analyses to accurately identify the location of the hyaluronidase gene.

The goal of this research is to utilize hybrid sequencing methods to accurately locate the hyaluronidase gene within S. equi subsp. zooepidemicus MTCC 3523. By accurately identifying the position of this gene, we can create targeted mutagenesis approaches to disrupt hyaluronidase expression, thus increasing HA production and encouraging the generation of high-molecular-weight HA. Along with gene knockout methods, this study seeks to deliver an extensive comparative genomic analysis of S. equi subsp. zooepidemicus MTCC 3523, examining its phylogenetic connections, genomic consistency, and potential metabolic pathways for large-scale HA production. This investigation will aid in optimizing microbial HA production, providing valuable insights into genetic alterations that can enhance HA biosynthesis, and laying the groundwork for future developments in biopharmaceutical and cosmetic fields.

MATERIAL AND METHODS

1.1. DNA Extraction and Quality Control

Genomic DNA was extracted from bacterial cells using the QiagenDNeasy Blood and Tissue Kit (Cat No. 69506). The bacterial cell pellet was dissolved in lysozyme (10 mg/mL; Cat No. L6876) and incubated at 37°C for 30 minutes. For cell lysis, AL buffer and Proteinase K were used at 56°C for a duration of 2 hours. RNase A treatment (50 mg/mL; Cat No. 2101076) was performed at 65°C for 30 minutes. The lysate was mixed well with half its volume of absolute ethanol and transferred to a DNeasy Mini Spin Column situated in a 2 mL collection tube. After centrifugation at 8,000 rpm, the flow-through was discarded. The subsequent wash steps were followed according to the manufacturer's instructions. DNA was eluted using 10 mMTris-HCl (pH 8.0). The concentration and purity of the genomic DNA were evaluated with a Nanodrop Spectrophotometer (Thermo Scientific 2000) and quantified using the QubitdsDNA HS Assay Kit (Cat No. Q32854).

1.2. Sanger Sequencing

To verify the purity of the bacterial strain, amplification of the 16S rRNA gene was conducted. PCR amplification utilized approximately 30 ng of genomic DNA as a template, along with 16S rDNA primers (27F and 1492R), and Takara Ex Taq Polymerase (Cat No. RR001A) in a total volume of 25 µL. The successful amplification of an ~1.5 kb fragment was confirmed through agarose gel electrophoresis. The purified PCR products were sequenced by Sanger sequencing. A BLAST analysis identified the sample MTCC3523 as Streptococcus sp. 16S ribosomal RNA, showing a sequence identity match of 99.77%.

1.3. Illumina Library Preparation and Sequencing

Library preparation was conducted utilizing the QIASeq FX DNA Library Preparation Kit (Cat No. 180475) in accordance with the manufacturer's guidelines. Around 50 ng of DNA quantified with Qubit was enzymatically fragmented, end-repaired, and A-tailed in a single-tube reaction using the FX Enzyme Mix. The ligation of adapters was carried out with index-incorporated Illumina adapters to create the sequencing libraries. The libraries underwent enrichment through six cycles of indexing PCR (starting with an initial denaturation at 98°C for 2 minutes, followed by 6 cycles of 98°C for 20 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, concluding with a final extension at 72°C for 1 minute). The amplified libraries underwent purification using Cytiva beads (Magbio, #AC-60050), were quantified with a Qubitfluorometer (Thermo Fisher Scientific, MA, USA), and were evaluated for fragment size

distribution using an Agilent 2200 TapeStation. The libraries were sequenced in paired-end mode on an IlluminaNovaSeq X Plus sequencer (Illumina, San Diego, USA) for 150 cycles.

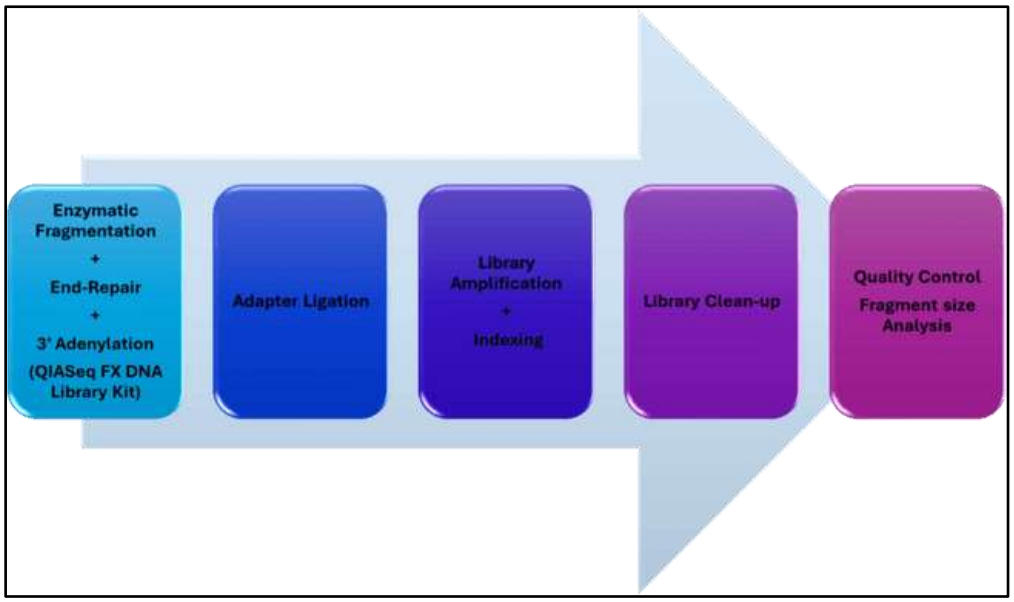
Enzymatic Fragmentation + End-Repair + 3' Adenylation (QIASeq FX DNA Library Kit)
Adapter Ligation
Library Amplification + Indexing
Library Clean-up
Quality Control Fragment size Analysis

Figure 1 The workflow diagram outlines the process of preparing DNA libraries using the QIASeq FX DNA Library Kit. The protocol initiates with enzymatic fragmentation, end-repair, and 3' adenylation of the DNA. Following this, adapter ligation is conducted, which allows for compatibility with sequencing. The subsequent stage includes library amplification and indexing, providing traceability for the samples. A library clean-up step follows to eliminate unwanted fragments and impurities. Lastly, a quality assessment and analysis of fragment size are performed to confirm the integrity of the library before moving forward with sequencing. This efficient method guarantees the creation of high-quality DNA libraries that are well-suited for next-generation sequencing applications.

1.4. Nanopore Library Preparation and Sequencing

Genomic DNA was subjected to end-polishing and A-tailing with the NEBNext Ultra II End Repair Kit (New England Biolabs, MA, USA). The prepared DNA was barcoded utilizing the Blunt TA Ligase Master Mix (M0367L), after which the barcoded samples were pooled in equimolar amounts. Sequencing adapters (SQK-LSK114.96) were ligated to double-stranded DNA fragments using NEB Quick T4 DNA Ligase (New England Biolabs, MA, USA). Following purification with AMPure XP beads, the final library was quantified using Qubit and then sequenced on the NanoporePromethION system (PromethION P24, Oxford Nanopore Technologies, UK) with a PromethION Flow Cell (FLO-PRO114M).

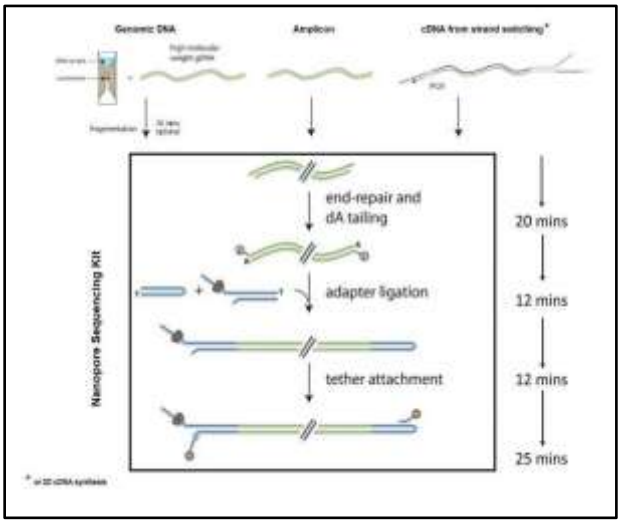

Figure 2 The Nanopore sequencing library preparation workflow is depicted schematically. This process starts with high-molecular-weight genomic DNA, cDNA from strand-switching PCR, or amplicons. The DNA is optionally fragmented for 30 minutes, followed by end-repair and dA-tailing for 20 minutes. Adapter ligation is then conducted

for 12 minutes, which facilitates the attachment of adapters compatible with sequencing. Next, tether attachment is completed in 12 minutes, ensuring effective loading onto the Nanopore sequencing flow cell. Finally, sequencing-ready molecules are introduced to the platform for real-time long-read sequencing that takes 25 minutes. This protocol allows for high-throughput sequencing of either native or amplified DNA and RNA molecules (Reprinted with permission from 26).

For the sequencing analysis, the bacterial strain Streptococcus equi subsp. zooepidemicus MTCC 3523 was used. The DNA sample extracted was processed using a barcoded sequencing method to ensure accurate tracking and high-quality sequencing results. The component sequence assigned to this sample was NB01 CACAAAGACACCGACAACTTTCTT, which was essential for identifying the sample during sequencing and data analysis.

DNA quantification was performed using a Qubitfluorometer, revealing a Qubit yield of 1.23 ng and a total DNA yield of 24.6 ng, indicating an adequate concentration of DNA for subsequent sequencing purposes. To enable indexing and multiplex sequencing, unique dual-index barcode sequences were designated for this sample. Barcode 1 index sequence was UDI042 TGTGGAACCG, and Barcode 2 index sequence was UDI042 CATTGACTCT. These barcodes guaranteed precise demultiplexing and accurate identification of sequencing reads related to S. equi subsp. zooepidemicus MTCC 3523, contributing to reliable and high-fidelity genome assembly.

Data Analysis

Approximately 1.7 million paired-end Illumina sequencing reads and around 0.3 GB of Nanopore long-read data were generated. The Illumina raw reads underwent quality filtering and adapter trimming utilizing TrimGalore (v0.4.01)(12), while the Nanopore reads were processed for adapter removal with Porechop (v0.2.32)(13). A hybrid genome assembly was conducted using Unicycler (v0.4.83)(14), which combined both short- and long-read data. The resulting draft assembly yielded an approximately 2.1 Mb genome comprising 11 contigs. Gene and protein predictions were carried out using Prokka (v1.14)(15), with functional annotations assigned via DIAMOND BlastP(16) against the UniProt bacterial database. Metabolic pathways were identified through the use of KAAS (KEGG Automatic Annotation Server)(17). Phylogenetic analysis based on DNA-DNA hybridization was performed using the Type (Strain) Genome Server (TYGS)(18). A circular genome comparison was executed with the reference strain Streptococcus equi subsp. zooepidemicus NCTC 7023 using BRIG (v0.95)(19), and genome synteny analysis was conducted with Mauve (v2.4.0)(20). A comprehensive bioinformatics workflow is illustrated in Figure 3.

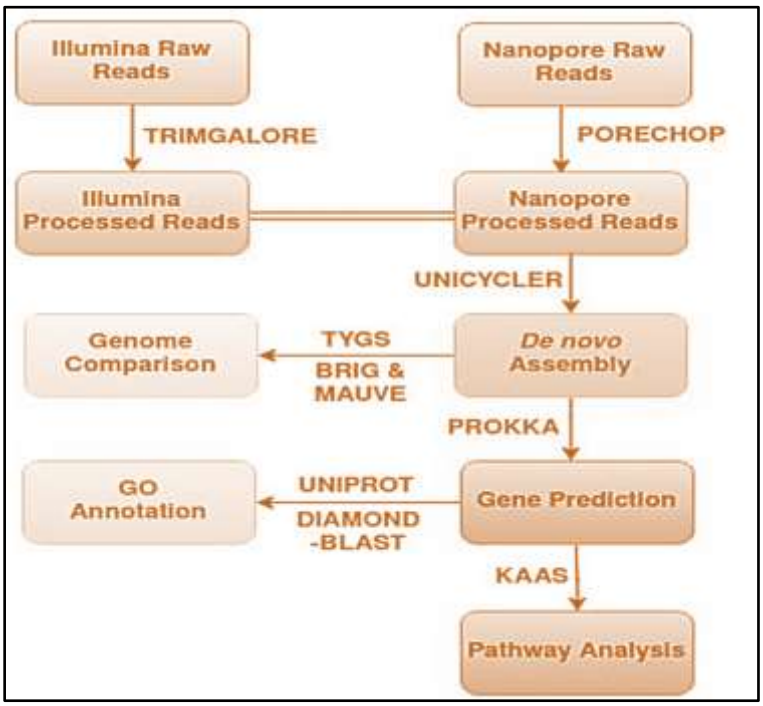

Figure 3A bioinformatics pipeline for the assembly, annotation, and functional analysis of Streptococcus equi subsp. zooepidemicus MTCC3523. Raw reads from Illumina and Nanopore are processed with TrimGalore and Porechop, respectively. The processed reads are then combined and assembled using Unicycler for de novo genome construction. Gene prediction is carried out using Prokka, followed by functional annotation with UNIPROT and DIAMOND-BLAST for Gene Ontology (GO) classification. Pathway analysis is conducted with KAAS, while genome comparisons are performed using TYGS, BRIG, and MAUVE.

RESULTS

DNA Quality Control

Quality assessment verified an optimal yield and concentration of DNA for both Illumina and Nanopore library preparations. In evaluating the quality control and sequencing of S. equi subsp. zooepidemicus MTCC 3523, various essential metrics were analyzed. The DNA concentration measured by Nanodrop was 15.5 ng/µL, with an absorbance ratio of 1.67 for 260/280 nm and 0.5 for 260/230 nm, reflecting moderate purity. The total volume of DNA used for sequencing was 25 µL, producing 387.5 ng of extracted DNA.

Additional quantification through Qubitfluorometry indicated a concentration of 4.52 ng/µL, with a total volume of 25 µL, leading to a final DNA yield of 113 ng according to Qubit measurements. The quality control assessment categorized the purity as sub-optimal; however, the overall DNA yield was considered acceptable for sequencing purposes.

For Illumina sequencing, a total of 1,703,972 reads were produced, with 1,669,374 reads successfully processed after quality filtering. This yielded a high sequencing coverage of 232X, ensuring solid genome assembly and subsequent analysis. These findings validate the suitability of the extracted DNA for whole-genome sequencing and facilitate high-confidence data interpretation.

Illumina Library QC and Data Demultiplexing

The sequencing library demonstrated an average fragment length of 451 bp, with an adequate concentration for sequencing. The data generated from the sequencing run was demultiplexed using Bcl2Fastq (v2.20), producing FastQ files based on distinct dual barcodes. The quality of the sequencing was evaluated using Commander software (Figure 4).

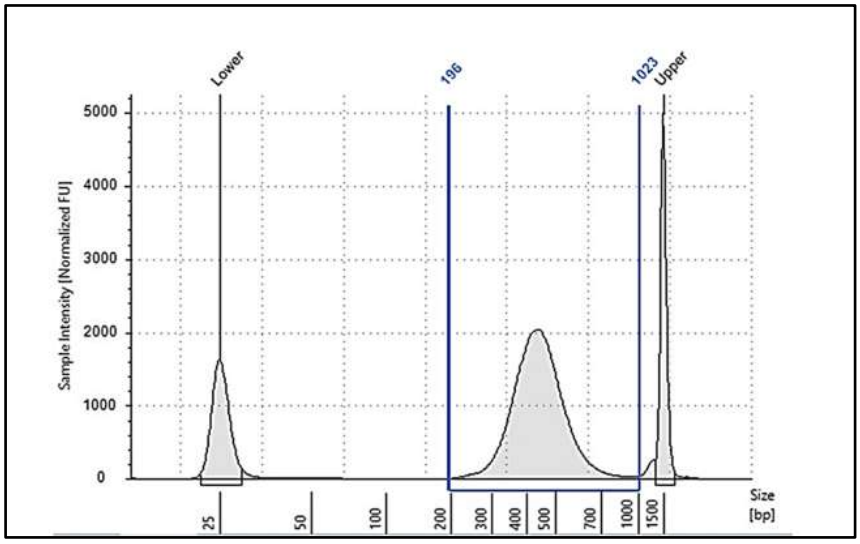

Figure 4 The electropherogram from the TapeStation for Streptococcus equi subsp. zooepidemicus MTCC3523 shows DNA fragment sizes between 196 bp and 1023 bp, with an average size of 451 bp. The DNA concentration is recorded at 21.0 ng/µL, resulting in a regional molarity of 74.7 nmol/L, which represents 96.43% of the total DNA content. Size standards used for precise fragment analysis are indicated by the lower and upper markers.

Primary Data Analysis

A total of around 1.7 million paired-end Illumina reads and 0.3 GB of long reads from Nanopore were collected. The Illumina raw reads underwent quality filtering using TrimGalore (removing those with a Phred score below 30 and lengths shorter than 20 bp), while the Nanopore reads were trimmed with Porechop.

SECONDARY DATA ANALYSIS

De Novo Hybrid Assembly

Genome assembly was conducted using Unicycler, combining both Illumina and Nanopore sequencing reads. The resulting draft genome comprised 11 contigs, totaling approximately 2.1 Mb in size. In our sequencing analysis of S. equi subsp. zooepidemicus MTCC 3523, we generated raw reads to support thorough genome assembly and annotation. A total of 836,174 reads were generated, contributing to an overall base count of 0.34 GB. The sequencing process yielded an average read length of 401.1 base pairs, with a mean quality score of 12, which ensured a high percentage of

reliable sequences. Importantly, 99.90% of the reads had a quality score greater than Q7, reflecting minimal sequencing errors.

The estimated sequencing coverage was 152.47X, which provided adequate depth for accurate genome reconstruction. The assembly process produced 11 contigs, with the largest measuring 1.6 MB and the smallest at 119 base pairs. The overall length of the assembled contigs reached 2.1 MB, resulting in a nearly complete draft genome of the bacterium. A key quality measure of the assembly, the N50 value, was found to be 1.6 MB, indicating a well-structured genome with extended contiguous sequences. The assembled genome showed no non-ATGC characters, confirming the integrity and completeness of the sequence data. Additionally, the assembly featured five contigs longer than 10 Kbp, with one contig exceeding 1 Mbp, underscoring the effectiveness of the sequencing approach in capturing large genomic portions.

These sequencing findings lay a robust groundwork for additional genomic investigations, including functional annotation, metabolic pathway reconstruction, and comparative analyses with closely related strains.

Gene Prediction and GO Annotation

Gene prediction was conducted with Prokka, resulting in the identification of 2,008 predicted proteins. Gene ontology (GO) annotation was performed using DIAMOND BlastP against the UniProt bacterial database, with a minimum identity threshold of 30% set for significant hits (Figure 5).

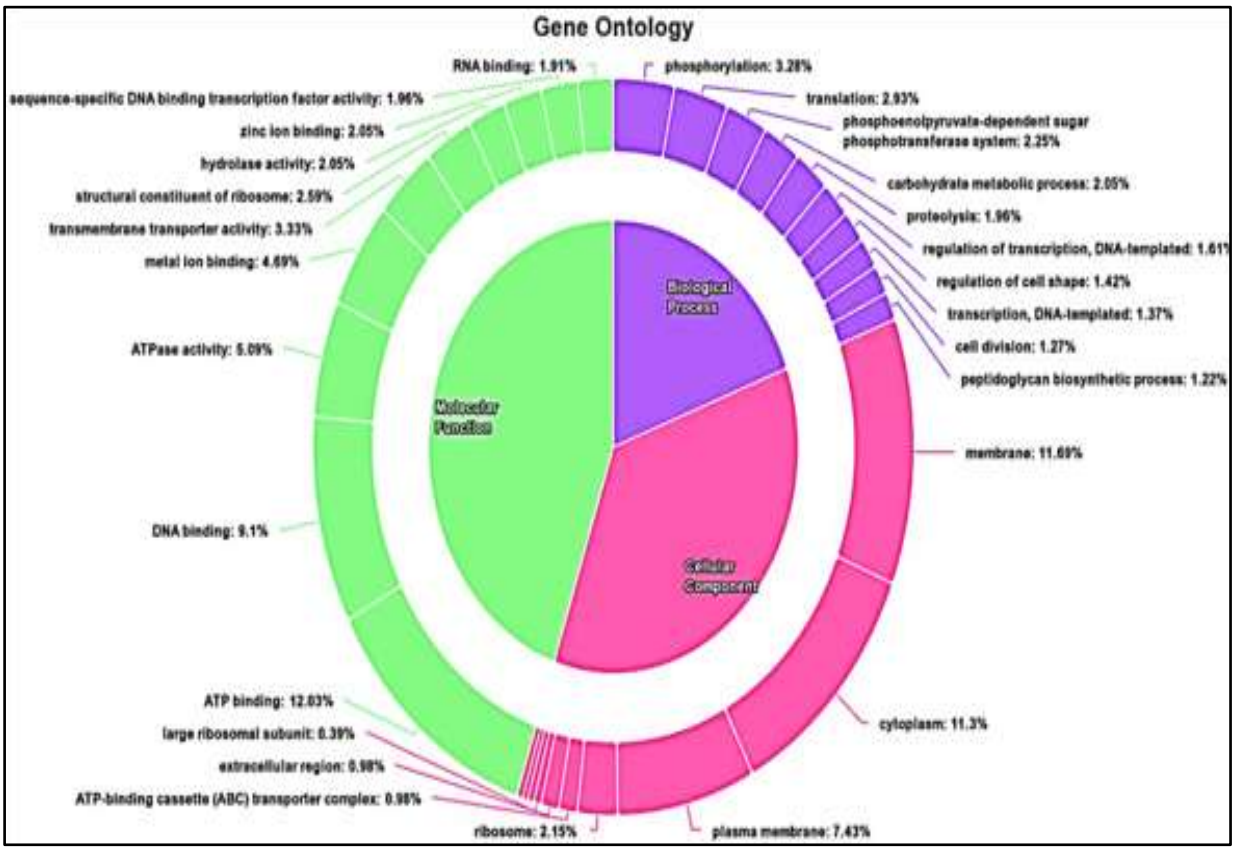

Figure 5 The Gene Ontology (GO) classification for annotated genes in Streptococcus equi subsp. zooepidemicus is presented MTCC3523. The circular diagram illustrates the distribution of genes among three primary GO categories: Molecular Function (green), Biological Process (purple), and Cellular Component (pink). Notable molecular functions include ATP binding (12.03%) and DNA binding (9.1%), while key biological processes encompass phosphorylation (3.28%) and translation (2.93%). The cellular component category underscores

Pathway Analysis

Pathway annotation was performed utilizing KAAS. The forecasted genes were assigned to KEGG pathways, referencing datasets such as Streptococcus equi subsp. zooepidemicus MGCS10565 and H70. Figure 6 presents the leading 10 KEGG pathway entries.

International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 13, Issue 5, May-2025, Available online at: www.ijaresm.com

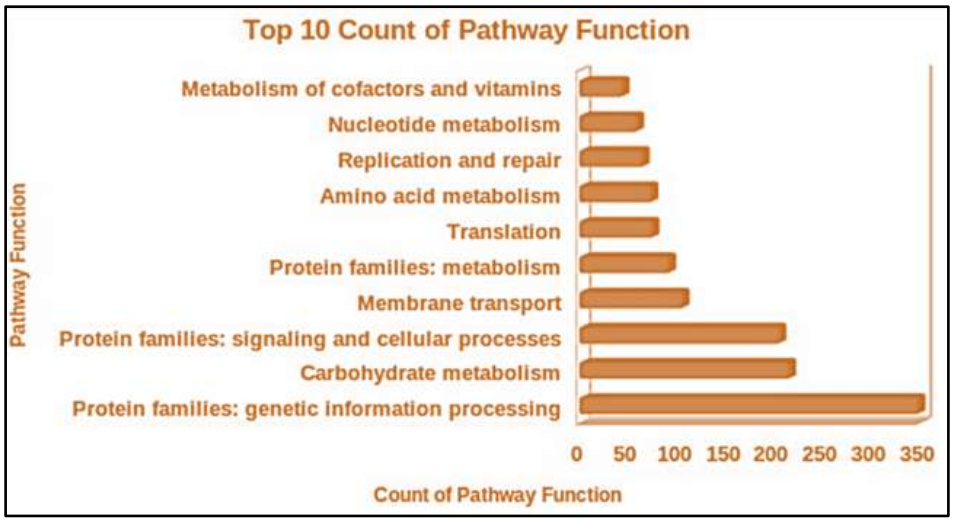

Figure 6 Top 10 pathway functions identified through gene annotation in Streptococcus equi subsp. zooepidemicus MTCC3523. The bar chart illustrates the frequency of pathway functions grouped into primary metabolic and cellular processes. The most predominant category is "Protein families: genetic information processing," with "Carbohydrate metabolism" and "Protein families: signaling and cellular processes" following closely, highlighting the main biological pathways active in this strain.

DNA-DNA Hybridization Analysis

A phylogenetic analysis was performed using TYGS. Whole-genome comparisons revealed that sample MTCC3523 shares a close relationship with Streptococcus equi subsp. zooepidemicus NCTC 7023, exhibiting a DNA-DNA hybridization (dDDH) sequence identity of 100%. The phylogenetic tree based on whole genomes is illustrated in Figure 7.

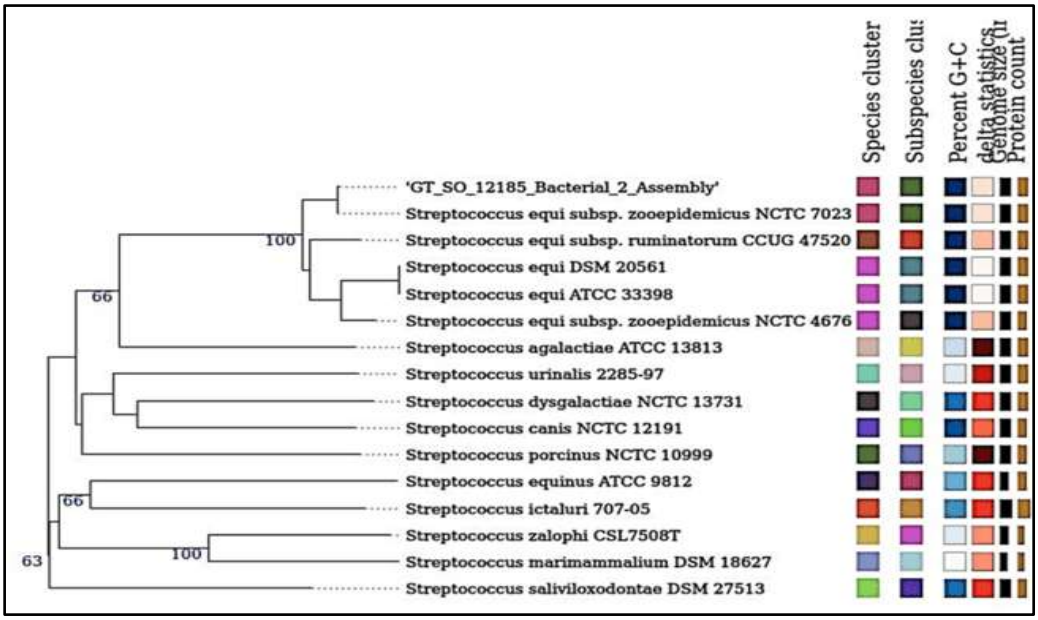

Figure 7 Phylogenetic tree representing various species of Streptococcus, including Streptococcus equi subsp. zooepidemicus and related groups. This tree was created based on genomic sequencing data, with bootstrap values shown at significant branches. The annotations on the right offer details about species groupings, classification of subspecies, GC content, and the number of proteins. The position of MTCC3523 within the S. equi cluster indicates its evolutionary relationship with other strains.

Comparative Genome Analysis

Genome comparisons in a circular format were conducted with BRIG, emphasizing the sequence similarities to S. equi subsp. zooepidemicus NCTC 7023. Synteny analysis carried out with Mauve revealed possible genomic rearrangements or recombination events (Figures 8a and 8b).

International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 13, Issue 5, May-2025, Available online at: www.ijaresm.com

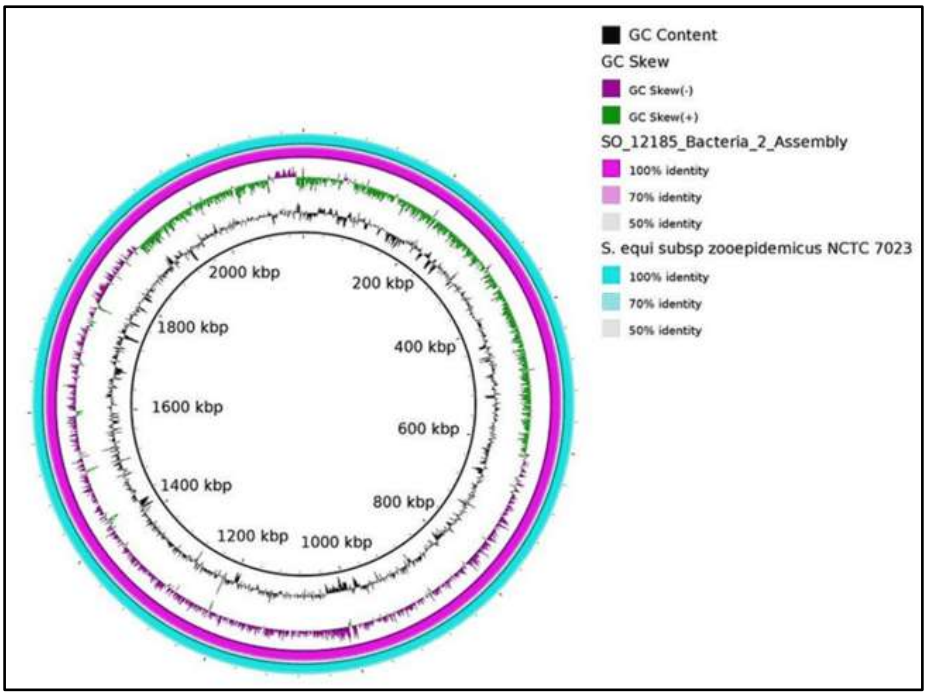

Figure 8 Comparative analysis of the circular genome of Streptococcus equi subsp. zooepidemicus MTCC3523 with that of S. equi subsp. zooepidemicus NCTC 7023 is presented. The innermost ring shows GC content (black), whereas the next ring indicates GC skew (with purple denoting negative skew and green representing positive skew). The outer rings illustrate the sequence identity between the assembled genome MTCC3523 and the reference genome (S. equi subsp. zooepidemicus NCTC 7023) at various identity thresholds (100%, 70%, and 50%). The high degree of sequence similarity indicates a close evolutionary relationship between the two strains.

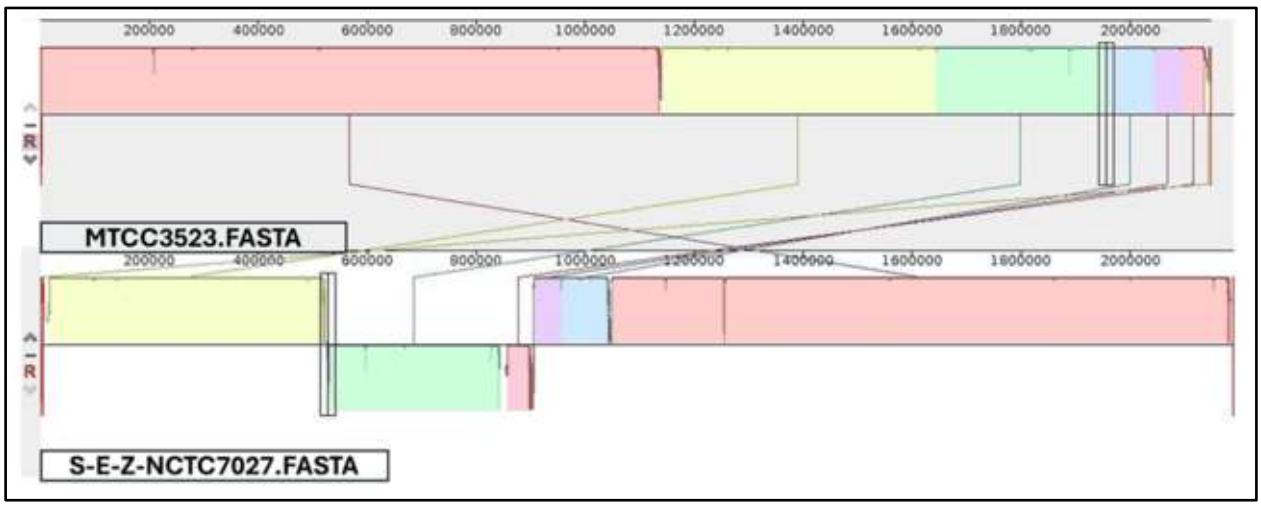

Figure 9 The whole-genome alignment of Streptococcus equi subsp. zooepidemicus MTCC3523 (MTCC3523.FASTA) and S. equi subsp. zooepidemicus NCTC 7027 (S-E-Z-NCTC7027.FASTA) was conducted using the MAUVE alignment tool. The colored blocks signify locally collinear sequence regions, where homologous segments are retained between the genomes. The existence of rearrangements, inversions, and gaps reflects structural variations between the two strains. Lines that connect the blocks indicate syntenic regions, while intersecting lines reveal genome rearrangements or inversions.

DISCUSSION

This research offers a thorough genomic examination of Streptococcus equi subsp. zooepidemicus by integrating Illumina short-read and Nanopore long-read sequencing methods. The hybrid assembly technique resulted in a high-quality draft genome with improved contiguity and completeness, aiding in downstream functional annotation and comparative genomic studies.

Oxford Nanopore sequencing is a fourth-generation technology that allows for the immediate generation of long-read sequences. It employs a nanopore-based electrical sensing method, where single-stranded DNA threads through a protein nanopore in a membrane, and the resulting disruption in ionic current is measured to identify the nucleotide sequence. This technology has numerous advantages, such as producing ultra-long reads (over 100 kb), sequencing in real time, and offering portability. These characteristics make it particularly advantageous for addressing complex genomic regions, structural variations, and hybrid genome assemblies. However, Nanopore sequencing tends to have a higher base-calling error rate compared to short-read technologies, making hybrid strategies essential for precise genome reconstruction.

Illumina sequencing, unlike other sequencing methods, is a second-generation technology that delivers high-throughput, short-read sequences with a high degree of accuracy. It operates on a sequencing-by-synthesis principle, wherein fluorescently labeled nucleotides are added to DNA strands in a cyclic detection process. This technology is extensively utilized for whole-genome sequencing, transcriptomics, and metagenomics thanks to its low error rate and cost-effectiveness. Nevertheless, its limited read lengths (usually between 150 and 300 bp) constrain its capacity to accurately resolve repetitive sequences and structural variations. The combination of Illumina and Nanopore sequencing leverages the advantages of both approaches—achieving high accuracy from short reads while enhancing genome assembly with long reads.

Hybrid de novo assembly proves especially useful for bacterial genomes where the precise sequence is not known. By merging short, precise Illumina reads with longer, more error-prone Nanopore reads, these hybrid assembly techniques facilitate the accurate reconstruction of complete bacterial genomes. This method is crucial for detecting novel bacterial strains, analyzing uncultured microorganisms, and clarifying intricate genomic structures such as repeat regions, plasmids, and mobile genetic elements. In instances where the bacterial genome hasn't been sequenced previously, hybrid assembly offers a dependable framework for constructing contiguous genomes that are more reflective of their true genetic architecture.

The effective extraction and assessment of genomic DNA resulted in the creation of high-quality sequencing data. The sequencing metrics demonstrated optimal library preparation and performance, generating sufficient coverage to support hybrid assembly. Illumina sequencing produced approximately 1.7 million paired-end reads, while Nanopore sequencing provided around 0.3 GB of long-read data. The amalgamation of these datasets within Unicycler resulted in an assembled genome of roughly 2.1 Mb consisting of 11 contigs, which marked a notable enhancement in assembly quality compared to approaches relying solely on short reads.

Functional annotation identified 2,008 protein-coding genes, underscoring essential metabolic and cellular functions. Gene ontology (GO) and KEGG pathway analyses showed the presence of genes linked to carbohydrate metabolism, virulence factors, and antibiotic resistance, which could influence bacterial adaptability and pathogenicity. Additionally, the KEGG pathway analysis reaffirmed the similarities to S. equi subsp. zooepidemicus reference strains, supporting the classification of the isolate.

The whole-genome phylogenetic analysis and DNA-DNA hybridization indicated that the bacterial strain is closely related to S. equi subsp. zooepidemicus NCTC 7023, exhibiting a perfect sequence identity of 100%. Comparative genome analyses utilizing BRIG and Mauve uncovered conserved genomic regions and synteny with the reference strain, suggesting limited genomic rearrangements. This observation provides valuable insight into the genetic stability of this subspecies and its evolutionary ties within the Streptococcus equi lineage.

In addition to classifying taxa, the research underscores the functional capabilities of the genes identified. The existence of genes linked to antibiotic resistance mechanisms is especially important, as it implies a possible role in the survival and adaptability of bacteria under selective pressure. Conducting further functional validation of these genes could clarify their specific functions and their contributions to bacterial pathogenicity. Moreover, the discovery of genes connected to carbohydrate metabolism suggests that S. equi subsp. zooepidemicus can utilize various nutrient sources, which could be critical for its colonization and persistence in different host environments.

A significant industrial and biomedical property of S. equi subsp. zooepidemicus is its capability to synthesize hyaluronic acid (HA), a high-molecular-weight polysaccharide used in medicine(27), cosmetics, and biotechnology(21,22). The HA biosynthesis pathway in S. equi subsp. zooepidemicus is thoroughly characterized and involves the enzymes hyaluronan synthase (HasA), UDP-glucose dehydrogenase (HasB), and UDP-glucose pyrophosphorylase (HasC). These genes facilitate the polymerization of HA from precursor sugars, and their presence in the assembled genome supports the strain's ability to produce HA. Given the rising demand for microbial HA as a sustainable alternative to animal-derived sources, further examination of the regulatory mechanisms that control HA biosynthesis in S. equi subsp. zooepidemicus could yield valuable insights for industrial use.

Conversely, S. equi subsp. zooepidemicus also generates hyaluronidase, which is an enzyme that breaks down hyaluronic acid. Hyaluronidase operates by cleaving HA into smaller segments, affecting bacterial invasion, evasion of host immunity, and the formation of biofilms. The equilibrium between HA synthesis and degradation is vital for bacterial virulence and adaptability since excessive activity of hyaluronidase may result in diminished HA production, potentially restraining its industrial applications. In our study, we pinpointed the exact location of the hyaluronidase gene, allowing for targeted mutagenesis to reduce its activity, thereby increasing HA yield and supporting the formation of longer HA chains. Gaining a better understanding of the regulatory mechanisms that manage hyaluronidase expression in S. equi subsp. zooepidemicus is crucial for optimizing HA production while reducing unwanted degradation. Genetic alterations or targeted inhibition of hyaluronidase activity may offer a promising approach to boost microbial HA yields for commercial purposes.

The results of this research also aid in comprehending the evolutionary path of S. equi subsp. zooepidemicus. Comparative genomic studies with closely related strains suggest that horizontal gene transfer (HGT) and genomic rearrangements could have influenced the genome of this subspecies. This supports the idea that bacterial adaptation to various ecological niches is enhanced by genomic plasticity. Future research might delve deeper into the extent of HGT events and their functional implications in different bacterial populations.

In summary, this study effectively illustrates the use of hybrid sequencing technologies for characterizing bacterial genomes, providing a solid foundation for forthcoming investigations into the functional and comparative genomics of S. equi subsp. zooepidemicus. Future efforts may concentrate on further clarifying the functional roles of predicted genes, examining strain-specific differences, and exploring potential virulence factors that contribute to pathogenicity and host interactions. Additionally, whole-genome comparative studies with both pathogenic and commensal Streptococcus strains could lead to a deeper understanding of the genomic traits that delineate virulence and host specificity. Expanding the dataset with more bacterial isolates from a range of environmental sources would improve our comprehension of genomic diversity and evolutionary dynamics in this bacterial species.

CONCLUSION

This research emphasizes the benefits of utilizing hybrid whole-genome sequencing with both Illumina and Nanopore technologies for detailed genomic analysis. By integrating the high precision of Illumina sequencing with the long-read potential of Nanopore technology, we were able to create a more comprehensive and contiguous assembly of S. equi subsp. zooepidemicus MTCC 3523. The results offer greater understanding of bacterial functional genomics, especially regarding the regulation of HA biosynthesis, setting the stage for targeted genetic modifications aimed at increasing HA yield. This study's outcomes have important implications for industrial biotechnology, providing a solid basis for optimizing strains for pharmaceutical and cosmetic uses. Future research should concentrate on enhancing sequencing accuracy, advancing genome assembly tools, and investigating innovative bioengineering techniques for improving microbial strains.

COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments
We are truly grateful for the invaluable support from Genotypic Company(28). Our heartfelt thanks go to Professor Mehdi Totonchi for his scientific guidance and outstanding leadership throughout this project. We also appreciate ShafaghNewGene Medical Company for their funding and support, which enabled this research to take place.

Conflict of Interest Disclosure
There are no conflicts of interest to report.

Ethical Approval Statement
This project received approval from the Research Ethics Committee of Islamic Azad University, Tehran Medical Sciences University – Faculty of Pharmacy and Pharmaceutical Branches, under the ethical code IR.IAU.PS.REC.1402.621.

REFERENCES

[1] Mihai Pop, Genome assembly reborn: recent computational challenges, Briefings in Bioinformatics, Volume 10, Issue 4, July 2009, Pages 35366, https://doi.org/10.1093/bib/bbp026

[2] Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodama M, Nielsen FC. Whole genome sequencing in clinical practice. BMC Med Genomics. 2024 Jan 29;17(1):39. doi: 10.1186/s12920-024-01795-w. PMID: 38287327; PMCID: PMC10823711.

[3] Martins VB, Moreira TD, da Silva Júnior AH, Sayer C, de Mello JM, Immich AP. Brewing Industry By-products: An Innovative Alternative to Hyaluronic Acid Biosynthesis. Current Analytical Chemistry. 2025 Jan 6.
[4] Pugh J. The Current State of Nanopore Sequencing. Methods Mol Biol. 2023;2632:3-14. doi: 10.1007/978-1-0716-2996-3_1. PMID: 36781717.
[5] Sharon BM, Hulyalkar NV, Nguyen VH, Zimmern PE, Palmer KL, De Nisco NJ. Hybrid De Novo Genome Assembly for the Generation of Complete Genomes of Urinary Bacteria using Short- and Long-read Sequencing Technologies. J Vis Exp. 2021 Aug 20;(174). doi: 10.3791/62872. PMID: 34487123.
[6] Lerminiaux N, Fakharuddin K, Mulvey MR, Mataseje L. Do we still need Illumina sequencing data? Evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. Can J Microbiol. 2024 May 1;70(5):178-189. doi: 10.1139/cjm-2023-0175. Epub 2024 Feb 14. PMID: 38354391.
[7] Shukla P, Srivastava P, Mishra A. On the potential activity of hyaluronic acid as an antimicrobial agent: experimental and computational validations. Bioprocess and Biosystems Engineering. 2025 Jan;48(1):27-42.
[8] Contreras Mendoza J, Arriola Guevara E, Suarez Hernández LA, Toriz G, Guatemala-Morales GM, Corona-González RI. Evaluation of mango residues to produce hyaluronic acid by Streptococcus zooepidemicus. Folia Microbiologica. 2024 Aug;69(4):847-56.
[9] Harth ML, Furlan FF, Horta AC. Microbial hyaluronic acid production in the 21 century: a roadmap toward high production, tailored molecular weight. OBSERVATÓRIO DE LA ECONOMÍA LATINOAMERICANA. 2024 Mar 27;22(3):e3913-.
[10] Bakar A, Mukhriza TM. Microbial Hyaluronic Acid Production: A Comprehensive Review of Strategies, Challenges and Sustainable Approaches. JurnalSerambi Engineering. 2024 Jun 7;9(2):9138-50.
[11] Abdullah Thaidi NI, Mohamad R, Wasoh H, Kapri MR, Ghazali AB, Tan JS, Rios-Solis L, Halim M. Development of in situ product recovery (ISPR) system using amberlite IRA67 for enhanced biosynthesis of hyaluronic acid by streptococcus zooepidemicus. Life. 2023 Feb 16;13(2):558.
[12] Babraham Bioinformatics. (n.d.). Trim Galore! Available at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
[13] Wick, R. (n.d.). Porechop. Available at: https://github.com/rrwick/Porechop.
[14] Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoSComput Biol. 2017;13(6):e1005595. Published 2017 Jun doi:10.1371/journal.pcbi.1005595
[15] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. Epub 2014 Mar 18. PMID: 24642063
[16] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12(1):59-60. doi: 10.1038/nmeth.3176. Epub 2014 Nov 17. PMID: 25402007.
[17] Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W182-5. doi: 10.1093/nar/gkm321. Epub 2007 May 25. PMID: 17526522; PMCID: PMC1933193.
[18] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-ofthe-art genome-based taxonomy. Nat Commun. 2019;10(1):2182. Published 2019 May 16. doi:10.1038/s41467-019-10210-3
[19] Alikhan, NF., Petty, N.K., Ben Zakour, N.L. et al. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402 (2011). https://doi.org/10.1186/1471-2164-12-402
[20] Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome research, 14(7), 1394–1403. https://doi.org/10.1101/gr.2289704
[21] Beheshtizadeh N, Mohammadzadeh M, Mostafavi M, Seraji AA, Ranjbar FE, Tabatabaei SZ, Ghafelehbashi R, Afzali M, Lolasi F. Improving hemocompatibility in tissue-engineered products employing heparin-loaded nanoplatforms. Pharmacological Research. 2024 Jun 19:107260.
[22] Beheshtizadeh N, Amiri Z, Tabatabaei SZ, Seraji AA, Gharibshahian M, Nadi A, Saeinasab M, Sefat F, Kolahi Azar H. Boosting antitumor efficacy using docetaxel-loaded nanoplatforms: from cancer therapy to regenerative medicine approaches. Journal of Translational Medicine. 2024 May 30;22(1):520.
[23] Heikema AP, Horst-Kreft D, Boers SA, Jansen R, Hiltemann SD, de Koning W, Kraaij R, de Ridder MA, van Houten CB, Bont LJ, Stubbs AP. Comparison of illumina versus nanopore 16S rRNA gene sequencing of the human nasal microbiota. Genes. 2020 Sep 21;11(9):1105.
[24] Stefan CP, Hall AT, Graham AS, Minogue TD. Comparison of illumina and Oxford nanopore sequencing technologies for pathogen detection from clinical matrices using molecular inversion probes. The Journal of Molecular Diagnostics. 2022 Apr 1;24(4):395-405.
[25] Cook R, Brown N, Rihtman B, Michniewski S, Redgwell T, Clokie M, Stekel DJ, Chen Y, Scanlan DJ, Hobman JL, Nelson A. The long and short of it: Benchmarking viromics using Illumina, Nanopore and PacBio sequencing technologies. Microbial genomics. 2024 Feb 29;10(2):001198.

[26] Cummings PJ, Olszewicz J, Obom KM. Nanopore DNA sequencing for metagenomic soil analysis. Journal of Visualized Experiments: JoVE. 2017 Dec 14(130):55979.

[27] Tabatabaei, S. Z., Motevaseli, E., SamieeZafarghandi, Z., Tabatabaei, S. A. Comparison of the Overexpression of HOTAIR lncRNAs and the Downregulation of HOXD10 Between Familial and Sporadic Coronary Artery Diseases. Iranian Heart Journal, 2021; 22(4): 127-134.

[28] Tabatabaei, S. Z., Motevaseli, E., SamieeZafarghandi, Z., Tabatabaei, S. A. Comparison of the Overexpression of HOTAIR lncRNAs and the Downregulation of HOXD10 Between Familial and Sporadic Coronary Artery Diseases. Iranian Heart Journal, 2021; 22(4): 127-134.

[29] Genotypic Technology Pvt. Ltd. (2023) 'NGS Sequencing and Genome informatics services', Available at: https://www.genotypic.co.in (Accessed: 12 October 2023).